# A COMPARISON OF SELF-REPORTED MEASURES OF PERCEIVED HEALTH AND FUNCTIONAL HEALTH IN AN ELDERLY POPULATION

Richard L. Lichtenstein, Ph.D. and J. William Thomas, Ph.D.

ABSTRACT: In studies of large elderly populations, two types of measures of physical health status, perceived health and functional health, are commonly used. Although they represent very different conceptions of health, these two types of measures appear often to be used interchangeably. In this paper, we examine changes over time in self-reported measures of perceived health and functional health for a sample of Medicare beneficiaries. By investigating the patterns of change in the two measures for different subgroups of the population, we are able to draw inferences about the appropriateness of each type of measure for specific administrative and/or research situations. The perceived health status measure appears suitable for descriptive studies of the health of elderly populations, while the greater stability of functional health makes this type of measure generally more appropriate in studies investigating relationships between an individual's physical health status and subsequent behavior.

## INTRODUCTION

German[1] has suggested that work in health status assessment for the elderly falls into two general categories: activities concerned with the direct delivery of services, and studies concerned with policy and program decisions. The former category would include, for example, determining the amount and level of nursing care a patient might require while institutionalized; determining the appropriate locus of care for a patient, such as home, hospital or skilled or basic care nursing home; and

213

evaluating progress of recovering patients. The latter category might include forecasting the use and costs of hospital and other health services, evaluating efficiency or effectiveness of specific clinical or social interventions, and monitoring health status changes over time in elderly populations.

A number of different measures of health status are available for these purposes.[2] Some are based on detailed clinical examinations, others on objectively observable behaviors, while still others rely on individuals' own assessments of their general health and well-being. The appropriate type of measure for a particular situation depends, as Ware et al.[3] have noted, on the purpose for which the measure is to be used. Because of practical issues involved in data collection, such as the feasibility and cost of obtaining accurate data, studies concerned with program and policy issues, typically involving work with large and possibly diverse populations, require instruments that can be self-administered or administered by nonprofessionals.[1]

Measures of physical health status which can be administered accurately, easily and inexpensively and which have been used frequently in studies of elderly populations include: those based on respondents' wholly subjective appraisal of their health status, so-called perceived health status measures, and those that relate to functional status, based on sets of activities respondents report they can, or cannot, do without assistance (or with assistance). Although such measures also are based on respondent's subjective assessments, these measures are viewed as more "objective" in that the behavior can be externally verified. A measure of the first type would be the often used single question, "Compared to other persons your age, would you say your health is excellent, good, fair, or poor?" Measures of the second type would include the Functional Limitation Battery developed by the Rand Corporation[1] and utilized by the National Center for Health Statistics in the National Medical Care Utilization and Expenditure Survey,[5] Katz's Activities of Daily Living (ADL) scale,[6] and Rosow and Breslau's Instrumental Activities of Daily Living (IADL) scale.[7]

In addition to the feasibility of data collection, a measure's validity in the specific context of the study being planned is recognized as an important consideration when choosing a health status measure. As Ware et al. note, the measure should contain information relevant to the particular aspect(s) of health status needed for the study and the planned analyses. However, they also observe that little work has been done to date investigating the validity characteristics of health status measures, and for many studies information needed for selecting one specific measure

over another is not yet available.[3] Thus, in research studies concerned with the elderly's utilization of hospital, physician, home care and other health services, perceived health status measures have been used by some,[8-10] functional measures by others[11-13] and both types of health status measures by still others.[14-17] Perceived measures have been used in efforts to predict mortality,[18] and functional measures to investigate active life expectancy.[19] Both types of measures have been used in investigations of health outcomes[20] and to describe the health of elderly populations.[21,22]

Perceived health status and functional health status represent very different conceptions which, though related statistically[17,23] clearly do not coincide. Perceived health represents the individual's evaluation of his or her total health compared to others, while functional health is a more narrowly focused assessment of physical functioning. Are such conceptual differences of practical importance; i.e., do they manifest themselves in ways that make one type of measure more appropriate than the other in specific situations? Are perceived measures, for example, more suitable than functional measures for studies explaining or predicting health services utilization patterns for the elderly? Are the two types of measures equally suitable for describing the health of populations?

The purpose of this paper is to explore characteristics of perceived and functional health status measures. Using data on a random sample of Medicare beneficiaries in Michigan, we trace changes over time in a perceived measure and a functional measure, and we look both at how health status changes for different age groups and at the degree to which measured health status is influenced by illnesses involving use of hospital and other health services. By examining such changes for different subgroups of the population, we are able to draw inferences about the construct validity of the measures themselves. These conclusions, we believe, may be useful for identifying which of the two types of measures is more appropriate in specific administrative and/or research situations.

## METHODS

Data for this analysis were originally collected for a study investigating the use of health status measures in adjusting Medicare capitation rates of HMOs. The subjects for this analysis were 1,616 Medicare beneficiaries from the State of Michigan who responded to two mail surveys conducted approximately one year apart, from October-December, 1982 and in January, 1984. The original sample consisted of 3,000 beneficiaries randomly selected from the Health

**TABLE 1**

Distribution of Scores on Perceived Health Scale for
Years One and Two (N = 1616)

| Response | % of Respondents | |
|---|---|---|
| | Year One | Year Two |
| (1) Excellent | 14.9 | 13.2 |
| (2) Good | 44.4 | 43.8 |
| (3) Fair | 33.2 | 35.3 |
| (4) Poor | 7.6 | 7.7 |
| | Mean = 2.33 | 2.37 |
| | SD = .820 | .808 |

Insurance Beneficiary State Tape (BEST) for Michigan, which listed all of the approximately 1 million beneficiaries in the state. A total of 2,123 completed questionnaires was returned during the first survey. After adjusting for beneficiaries who had died, moved or who could not be reached by mail or telephone follow-up interviews, the response rate to the first questionnaire was 81.2 percent.

A shorter survey instrument was mailed the following year to respondents in the first survey. The primary purpose of this questionnaire was to identify beneficiaries who had been hospitalized or received substantial health services outside of Michigan. A total of 1,683 (79%) of these second questionnaires were returned. Of those failing to respond, 85 had died, 30 could not be located, ten were functionally unable to respond and one refused to participate. After reviewing the returned questionnaires, fifty respondents were dropped due to out of state utilization (i.e. any hospital admission or three or more physician visits) and seventeen provided incomplete answers, leaving 1,616 individuals in the study. The following demographic characteristics, obtained in the first survey, describe the subjects: sex: 40.1% male; mean age in 1983: 75.8 (first quartile: 71 years; median: 75; third quartile: 79); level of education: 40% finished ninth grade, 80% finished twelfth grade; marital status: 55% married, 38% widowed; income: 27% earned less than $400 per month, 67% earned $800 or less per month; living situation: 33% living alone.

Two measures of beneficiary health status were developed from items included in both survey instruments. The first was a measure of the individual's perceived health status, derived from responses to the question, "Compared to other persons your age, would you rate your health as excellent, good, fair or poor?" The responses to this question were scored 1–4 respectively, and each respondent received a Year One and Year Two perceived health score. The distribution of scores in each year is presented in Table 1.

The other health status measure used in this analysis is an indicator of functional health. The second questionnaire included one item from Katz's Index of Activities of Daily Living[6] and three items from an Instrumental Activities of Daily Living scale[7]:

*Can you get dressed without help? (ADL)
*Can you walk up and down stairs to the second floor without help? (IADL)
*Can you walk half a mile without help? (IADL)
*Can you do heavy work around the house, like shoveling snow or washing walls, without help? (IADL)

These questions were also included in the first survey instrument, since it covered the complete set of items for both the ADL and IADL scales. Although responses to these self-reported functional status questions clearly do involve the perceptions of the respondents with respect to which activities they can do and which ones they need help with, we will call the scale composed of these items the "functional health" scale and will call the single item scale "perceived health." Functional health scores based on the four items were constructed for respondents for Year One and Year Two. Five levels of functional health scores, ranging from zero for those who could do none of the activities, to four for those who responded positively to all questions resulted. Scalogram Analysis of Year One and Year Two scores showed that this five-level functional health measure, like both the ADL and IADL, forms a Guttman scale; the Coefficient of Reproducibility was 0.99 for each year and the Coefficient of Scalability equaled 0.96 for Year One and 0.95 for Year Two. The top half of Table 2 shows the distributions of the five-level functional health scores for the two years. Since the major purpose of this analysis, however, was to compare changes in these two health status measures over time and since it would not be appropriate to compare changes in a four-point scale with those in a five-point scale, the two lowest levels of functioning in the functional health scale were collapsed. Because only 2.6 percent of cases were at the lowest scale level in Year One and only 4% were at the level in Year Two, we do not believe that combining these two levels causes any untoward effects. The distributions of the four-point functional health scale in Year One and Year Two are presented in the bottom half of Table 2.

Data on utilization of health services and Medicare expenditures were obtained directly from the Medicare intermediary/carrier in Michigan. Claims data on Supplemental Medical Insurance (Part B) services and on Health Insurance (Part A) services for hospital, skilled nursing facility, home health and outpatient care were included. Due to the long (approximately six month) lag time generally encountered by Medicare in processing and recording claims data and to the need to make the results of our larger study[24] available in timely fashion, we were able to collect data for only six months after the initial survey. Thus, in the analysis reported here, claims for services rendered between January 1, 1983

## TABLE 2

### Distribution of Scores on the Four and Five Level
### Functional Health Status Measures, Years One and Two (N = 1616)

| Scale Level | | Number and Percent of People at Each Scale Level in the Five Point Scale | | | |
|---|---|---|---|---|---|
| | | Year One | | Year Two | |
| | | # | % | # | % |
| Improving health status | 1 | 43 | 03 | 66 | 04 |
| | 2 | 142 | 09 | 176 | 11 |
| | 3 | 306 | 19 | 295 | 18 |
| | 4 | 502 | 31 | 438 | 27 |
| | 5 | 623 | 39 | 641 | 40 |

Mean Year One Five-level Functional Score = 3.94
SD = 1.08
Mean Year Two Five-level Functional Score = 3.87
SD = 1.17

| Scale Level | | Number and Percent of People at Each Scale Level in the Four Point Scale | | | |
|---|---|---|---|---|---|
| | | Year One | | Year Two | |
| | | # | % | # | % |
| Improving health status | 1 | 185 | 11 | 242 | 15 |
| | 2 | 306 | 19 | 295 | 18 |
| | 3 | 502 | 31 | 438 | 27 |
| | 4 | 623 | 39 | 641 | 40 |

Mean Year One Four-level Functional Score = 2.97
SD = 1.02
Mean Year Two Four-level Functional Score = 2.91
SD = 1.08

and June 30, 1983 are included, while claims for services rendered between July 1, 1983 and December 31, 1983 (the second survey was conducted in January, 1984) are not included. We do not believe that this data limitation affects the results presented in this paper.

## RESULTS

In examining changes in health status over the one year period of this study, both descriptive statistics and inferential statistics are pre-

sented for each of the health status measures considered. Subgroups defined by age, sex, utilization and expenditure characteristics of respondents have been created to enable analysis of these data across these important categories. Frequency distributions identifying percentages of respondents in the various subgroups whose scores remained the same over the year as well as those whose scores increased and decreased are presented first. Then, the intraclass correlation coefficient ($R_I$), an inferential statistic that has been used widely to test interrater reliability [25,26] is used here to measure the within-subject reliability of health status scores, i.e. the stability of an individual's scores over time.

Frequency distributions of changes in the health status scores for the two measures are presented in Table 3 and 4. These tables not only enable the identification of subgroups that experienced shifts in functional or perceived health, but they also permit the examination of the direction and magnitude of such shifts. Thus, Table 3 indicates that for the aggregate sample 70 percent of respondents had the same functional score in Years One and Two, while 11 percent experienced an increase in scores (improved health status) of one point, and nearly 2% experienced increased scores of two or more points. By comparison, 14% of the aggregate sample experienced a one-point decline in functional health score over the year, and declines of two or more points were reported by approximately 3% of all subjects.

Changes in functional health status for the specific subgroups are, as shown in Table 3, generally consistent with expectations. Persons in the lower risk groups, such as the "young" elderly (those 65–74 years of age) and those using fewer health services and incurring lower expenditures, were more likely than others to show no change in functional health status level. In the older age group and the groups experiencing higher rates of health services use and costs, relatively fewer persons remained at the same functional health level one year later and relatively more experienced declines in functional level. Among persons 85 years and older, 23% experienced declines in functional health status compared to only 17% for the aggregate sample. Among high health services users, 37% of persons with two or more hospital episodes during January-June 1983 experienced some decline in functional health, and declines were experienced by 33 percent of those having six-month expenditures of $5,000 or more.

Focusing on persons who reported 2+ point improvements in functional health status over the year, the patterns are more surprising. Subgroups with the greatest percentages in this category are the oldest age group, the highest hospital admission rate group and the highest expenditure group. From 4% to 6% of each of these high risk groups

## TABLE 3

### Distribution of Change in Functional Health Status Score Between Year One and Year Two for Various Subgroups of the Sample

Percent With Indicated Change in Score

| Change in Score | Aggregate Sample | Age | | | Sex | | Admissions | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 65–74 | 74–84 | 85+ | Male | Female | Zero | One | Two or More |
| Improved +2 or more | 1.8 | 3.2 | 1.0 | 4.0 | 1.7 | 1.9 | 1.6 | 3.0 | 4.8 |
| +1 | 11.0 | 9.9 | 12.6 | 10.3 | 8.9 | 12.4 | 11.9 | 5.4 | 2.4 |
| same | 70.0 | 72.2 | 69.2 | 62.6 | 73.0 | 67.9 | 70.5 | 68.7 | 56.1 |
| −1 | 14.2 | 13.9 | 14.0 | 16.7 | 13.0 | 15.0 | 13.7 | 15.7 | 26.8 |
| Declined −2 or more | 2.9 | 1.9 | 3.3 | 6.3 | 3.4 | 2.6 | 2.2 | 7.2 | 9.8 |

| Change in Score | Part B Services | | | Total Medicare Payments (Six Mo.) | | |
|---|---|---|---|---|---|---|
| | 0–14 | 15–29 | 30+ | $0–$999 | $1,000–$4,999 | $5,000+ |
| Improved +2 or more | 1.6 | 2.4 | 3.3 | 1.7 | 1.4 | 5.6 |
| +1 | 11.6 | 9.6 | 6.8 | 12.0 | 5.6 | 4.5 |
| even | 70.7 | 68.3 | 65.3 | 70.7 | 71.3 | 57.3 |
| −1 | 13.6 | 14.9 | 20.3 | 13.5 | 16.1 | 22.5 |
| Declined −2 or more | 2.5 | 4.8 | 4.2 | 2.2 | 5.6 | 10.1 |

Measure of Change in Functional Health (For Aggregate Sample)
Mean = −.053
SD = .693

# TABLE 4

## Distribution of Change in Perceived Health Score Between Year One and Year Two for Various Subgroups of the Sample

### Percent With Indicated Change in Score

| Change in Score | Aggregate Sample | Age 65–74 | Age 74–84 | Age 85+ | Sex Male | Sex Female | Zero | One | Two or More |
|---|---|---|---|---|---|---|---|---|---|
| Improved −2 or more | 0.8 | 0.6 | 0.6 | 3.4 | .8 | 0.9 | 0.8 | 1.8 | 0 |
| −1 | 15.2 | 15.3 | 15.1 | 15.5 | 14.7 | 15.6 | 15.7 | 10.9 | 17.5 |
| same | 64.1 | 65.3 | 63.5 | 60.3 | 63.7 | 64.3 | 64.9 | 57.6 | 60.0 |
| +1 | 18.7 | 18.2 | 18.9 | 20.1 | 19.7 | 18.0 | 17.7 | 26.7 | 20.0 |
| Declined +2 or more | 1.2 | 0.6 | 2.0 | 0.6 | 1.1 | 1.2 | 1.0 | 3.0 | 2.5 |

(Admissions: Zero, One, Two or More)

| Change in Score | Part B Services 0–14 | 15–29 | 30+ | Total Medicare Payments (Six Mo.) $0–$999 | $1,000–$4,999 | $5,000+ |
|---|---|---|---|---|---|---|
| Improved −2 or more | 1.0 | 0 | 0.8 | .8 | .7 | 2.2 |
| −1 | 16.7 | 7.3 | 12.7 | 15.7 | 13.4 | 11.4 |
| even | 63.2 | 69.9 | 63.6 | 65.0 | 59.2 | 58.0 |
| +1 | 18.1 | 12.4 | 19.5 | 17.6 | 24.6 | 25.0 |
| Declined +2 or more | 1.0 | 1.5 | 3.4 | 1.0 | 2.1 | 3.4 |

Measure of Change in Perceived Health (For Aggregate Sample)

Mean = .041

SD = .657

showed functional health status improvements of 2 points or more, compared to less than 2% of the aggregate sample experiencing this degree of improvement. For these groups, it is likely that higher percentages of persons experienced temporarily disabling health events prior to the initial survey, and that the improvements in functional health observed a year later represent either regression to the mean or the salutary effects of health care.

Unlike functional health status, the changes in perceived health scores shown in Table 4 appear to follow no consistent pattern. For the aggregate sample, the change distribution with the perceived measure is similar to that of the functional health measure, but with a smaller percentage, 64%, of the sample experiencing no change in perceived health over the year compared to 70% showing no change in functional health status. However, when the age and use/expenditure groups are examined, the distributional patterns for the two health status measures appear quite different. Table 3 shows that the percentage of respondents experiencing no change in functional health status is relatively high at about 72% in the youngest age group and drops off steadily reaching only 63% for the 85 year and older group. Also, for both the hospital-admission and total-Medicare-payments subgroups, the percentages of those remaining at the same functional health status level over the year is similar for each of the two lower categories (over 70%), but falls to below 57% in the highest use and highest expenditure categories. With the perceived measure, however, there is less difference among the age categories in the percentage of respondents at the same health status level of one year later. For the hospital-admissions categories, a somewhat smaller percentage of those with one admission remains at the same health status level compared to those with no admissions, but this percentage increases slightly for the group with two or more admissions. For the total-Medicare-payments subgroups, 65% of those with no payments remain at the same health status level, compared to 58% to 59% of those in the two higher expenditure categories.

In terms of percentage of respondents experiencing decreases or increases in health status over the year, the patterns are quite different for the two types of measures. The percentages of respondents showing declines in functional health status increase directly with the number of hospital admissions, ranging from 15.9 for those with no admission to 23 for those with one and 37 for respondents with two or more. For perceived health, no such relationship is apparent; 18.7% of those with no hospital admissions exhibit a health status decrease compared to 30% of those with one admission and 22.5 of those with more than one. Similar

differences between the two measures are observed for the age and the total-Medicare-payments groups.

In addition to examining percentage change distributions (Table 3 and 4), it is also interesting to view the issue of stability in measured health status for individuals from a scale reliability perspective using the intraclass correlation coefficient $(R_1)$. Although several different types of intraclass correlation coefficients have been discussed in the literature,[27] the one used here is based on a one-way analysis of variance (ANOVA) model using the individual rater (subject) as the unit of analysis. According to Bartko:[27]

> The one-way ANOVA intraclass correlation is given by:
>
> $$R_1 = (MSB - MSW)/[MSB + (C-1) MSW]$$
>
> Where $R_1$ = intraclass correlation, MSB = mean square between, MSW = Within = subjects variance, and C = Number of raters. The $R_1$ ranges from $-1$ (C-1) to 1.0. It is 1.0 when the within subjects variance is zero and the mean square between is greater than zero. A within variance of zero indicates identical ratings for a subject (i.e. where all the raters agree on the rating for that subject) and hence is consistent with a reliability of 1 or perfect agreement. A negative intraclass correlation is usually taken to be zero reliability. The $1 - R_1$ for intraclass correlation $\geq 0$ is interpreted as the percentage of variance due to the disagreement among the raters.

The $R_1$ values, and their 95 percent confidence limits, for the various subgroups identified previously are shown for both the perceived scores and the functional health scores in Table 5. According to Landis and Koch[25], the following "benchmarks" can be used to describe the "relative strength of agreement associated with $(R_1)$":

| Value of $R_1$ | Strength of Agreement |
|---|---|
| <0.00 | poor |
| 0.00-0.20 | slight |
| 0.21-0.40 | fair |
| 0.41-0.60 | moderate |
| 0.61-0.80 | substantial |
| 0.81-1.00 | almost perfect |

Using these guidelines we find that for most groups, $R_1$'s for both the perceived and functional health scores are in the substantial agreement range, and that the $R_1$'s for functional scores are nearly always

## TABLE 5

Stability of Functional Health and Perceived Health Scores
from Year One to Year Two as Measured by the Intraclass Correlation
Coefficient ($R_I$) for Various Subgroups of the Sample

| Sub-Group | Functional Health $R_I$** | Perceived Health $R_I$** | N |
|---|---|---|---|
| Aggregate Sample* | .78 ± .02 | .68 ± .03 | 1616 |
| Age: 65-74* | .78 ± .03 | .69 ± .04 | 798 |
| 75-84* | .78 ± .03 | .67 ± .04 | 643 |
| 85+ | .67 ± .08 | .63 ± .09 | 174 |
| Sex: Males* | .76 ± .03 | .68 ± .04 | 652 |
| Females* | .77 ± .03 | .66 ± .04 | 964 |
| Admissions: Zero* | .79 ± .02 | .68 ± .02 | 1409 |
| One* | .72 ± .07 | .53 ± .11 | 166 |
| Two or more | .50 ± .23 | .64 ± .18 | 41 |
| Part B Services:   0-14* | .79 ± .02 | .67 ± .03 | 1290 |
| 15-29 | .72 ± .07 | .69 ± .07 | 208 |
| 30+ | .66 ± .10 | .57 ± .12 | 118 |
| Total Medicare Payments (Six Months) | | | |
| $0-$999* | .79 ± .02 | .68 ± .03 | 1384 |
| $1,000-$4,999 | .79 ± .06 | .66 ± .09 | 143 |
| $5,000 or more | .51 ± .15 | .42 ± .17 | 89 |

*Difference in $R_I$ values is significant at $p \geq .05$.
**Also included with each $R_I$ value are approximate 95% confidence limits.

higher than those for perceived health. Many of the values for the functional scores approach almost perfect agreement while the $R_I$'s for the perceived scores are in the middle of the substantial agreement range. As the data in Table 5 show, for most subgroups the $R_I$ of the functional measure is significantly greater (at the 5 percent level) than that of the perceived health measure.* For subgroups where differences are not sta-

---

*Since techniques for testing differences between correlated intraclass correlation coefficients are not available, we have used 95 percent confidence intervals to determine whether, for specific subgroups, the functional measure $R_I$ is significantly different from the $R_I$ of the perceived health measure.[28] Where the confidence intervals of the two $R_I$'s that are being compared overlap the difference is judged not to be statistically significant at the 5 percent level. However, where the two confidence intervals do not overlap the difference is judged to be statistically significant.

tistically significant, the lack of significance is probably attributable to small sample size.

While the patterns of $R_1$ values for different subgroups is fairly similar for both health status measures, slightly different patterns are evident when looking at subgroups defined by number of hospital admissions and number of Part B claims claims. While $R_1$'s for the functional scores decline directly in relation to the amount of utilization experienced (i.e. the scores are less stable as beneficiaries used more care), the relationship between the $R_1$'s for perceived health and the experience of these different subgroups is not as clear. Perceived health scores are least stable for subjects with one admission while those with multiple admissions demonstrate stability similar to those with no admissions. This contrasts with the functional measure where the $R_1$ for the multiple admission subgroups is quite lower than the other two. Likewise, under Number of Part B claims, the middle category shows the highest $R_1$ value for perceived health whereas for functional health $R_1$'s decline as the amount of use experienced by the beneficiaries increases.

## DISCUSSION

Taken together, the data in Tables 3 through 5 indicate that both perceived and functional health measures remain quite stable over the course of one year, with functional health status exhibiting significantly greater stability than perceived health. Functional health status, though stable for the elderly population as a whole, declines for individuals who are in the "oldest old" category or who utilized large volumes of health care services during the year. Although some individuals, even in the high use/expenditure categories, experience improvements in functional health over the year, functional health status is far more likely to remain stable or to decline than it is to improve. Perceived health status appears to be less influenced than functional health status by illnesses requiring the use of high levels of health services. Distributions of changes in perceived health status for low use groups are quite similar to those for the high use group, and no pronounced decline in health status for those in the high use group is observed. Changes over the one-year period in perceived health are less skewed in the direction of poorer health status than are changes in functional health. Whereas the ratio of those who decline in functional health status to those who improve is as high as six-to-one for some subgroups, corresponding ratios for the perceived measures

are two-to-one or less. When interpreting these findings, obviously it would be preferable to gauge the stability of the alternative measures using some indicator of the respondents' "true" health status rather than simply their utilization or demographic characteristics. Unfortunately, true health is not directly observable or measurable, and so each of the two measures used in this study, and, for that matter, any other physical health status measure, is at best an imperfect indicator of some aspect(s) of an individual's true physical health state.[3] To consider such a measure "valid," in the construct validity sense then, the specific aspects of health to which the measure relates must be identified.* In examining the observed differences between perceived measures and functional measures, it is also useful to distinguish between the usual or "mean health" state of an individual over an extended period of time, and the individual's "episodic health," the individual's state of health at a particular point in time. Episodic health may change daily, declining with acute physical illness or swings in mood or emotional state and then improving; it varies around the individual's mean health. Although mean health may be expected to decline eventually in response to chronic illnesses or the disabling effects of severe trauma or acute illness, it will remain generally stable.[13] A valid health status measure can be expected to relate to one, or perhaps both, of these aspects.

With the perceived health measure used in this sutdy, an individual's score is likely to reflect at least three factors: (1) the objective influence of physical health, both episodic health at the time of the survey and the individual's longer term mean level of health; (2) the subjective influence of emotional status or mood, which, like physical health, may include a transient element, as well as the individual's outlook toward health and life in general which tends to remain fairly stable;[29,30] and (3) the person's perception of the health of those in the comparison group of those his or her age, i.e., the respondent's frame of reference. To the degree that individual ratings are more strongly influenced by episodic health and transient moods or emotions, perceived scores will exhibit short-term variability and apparent randomness. However, if longer term characteristics such as mean level of health and general outlook and

---

*As one would expect, both measures are related to mortality. However, it is interesting to note that the Year One functional measure is a slightly better predictor of mortality than the perceived measure. Among respondents who died during the year following the initial survey, 42.5% were in the lowest level of functional health whereas only 24.7% were in the lowest perceived level ("poor" health). In contrast, however, 16% of these respondents were in the highest level of functional health in Year One as opposed to only 5.5% who rated their health as "excellent". Nevertheless, in logistic regression analyses, controlling for age, sex, welfare status and institutional status, the functional health measure proved to be a better predictor of death than the perceived measure.

attitudes have the stronger influence, perceived scores will exhibit general stability.

The functional health measure differs from the perceived measure in at least two significant ways. First, it is probably less subjective since it is based, in part, on behavior which is observable. While emotional state may influence actual functioning or perceptions about ability to function for some individuals, for most respondents the report of ability to get dressed, to walk and to work will not vary with mood or outlook. Second, the measure uses an absolute, as opposed to relative, scale. Frame of reference is not an issue. Thus, unlike perceived health, an individual's functional score will generally be independent of the functional abilities of others.

Using this framework, we can now re-examine the data in Tables 3 through 5. First, we can note that the patterns shown in the tables suggest that both measures appear to be valid indicators of health, in that differences among groups are, with a few exceptions, consistent with expected differences in health. However, the percentage-change distributions in Table 3 and 4 and the $R_I$ values in Table 5 suggest that the perceived measure is more variable than the functional health measure. Except in those groups where changes in long-term mean health status might be expected, such as the 85-year-and-over age group and the two-or-more-hospitalizations group, the percentage of persons remaining at the same health status level after one year is consistently lower with the perceived measure than with the functional measure. Intraclass correlation coefficients also are consistently lower for the perceived measure than the functional measure. This may indicate that the perceived measure is more strongly related to episodic health and the functional measure to mean health.

For most people, mean health will remain stable over long periods of time and may be expected to deteriorate eventually with age or with the influence of chronic or significant acute illness. Data in Tables 3 and 5 for the functional health status measure appear to be consistent with this pattern. For the perceived measure, however, the percentage-change distributions and $R_I$ values are seen to vary only slightly with age and expenditures and inconsistently with use of services.

The weakness of these relationships is likely due to two factors. First, since the perceived measure is a relative scale, even if an individual's health deteriorates, resulting in more frequent illness, more health expenditures and even more hospital admissions, the person's view of his or her health relative to that of others of the same age may not change or, indeed, may improve if the health of the comparison group is perceived

as declining faster. Second, since episodic health is more likely than mean health to improve over time in an elderly population, some of the instability in the perceived measure may be due to actual improvements in health status. Whereas the changes in functional health shown in Table 3 tend to be negative, much of the change in perceived health is in the positive direction. Although some of these improvements may reflect better mean health levels, most probably reflect positive changes in episodic health. Thus, even a person hospitalized twice during the year for out-of-control diabetes could show an improved perceived health status score if her Year Two situation were felt to be better than that of Year One.

It is, of course, dangerous to generalize conclusions that are drawn from a study of two specific self-reported measures, and to suggest that other perceived and functional health measures, especially those that are not self-reported, might share the same characteristics. Therefore, in the following discussion we will restrict our comments concerning what we see as appropriate and inappropriate uses for these two types of health status measures only to measures that are self-reported and in which the perceived measure asks the respondents to compare their health to others their age.

As discussed by Ware et al.[3] and German,[1] both types of measures are of potential value for policy and program decisions. The easily obtained perceived health status scores appear to be particularly appropriate for describing differences among population groups when, for example, priorities are being established for health and social support programs or resources are being allocated to communities in which the elderly reside. For these applications, decisions are likely to be based on population mean values, and minor randomness in the measure at the level of the individual will be averaged out over the population.

In general, for studies concerned with investigating relationships between an individual's physical health status and specific health behaviors or events, a self-reported functional health status measure will be more suitable than a perceived measure since random variation or "noise" in the perceived measure may mask the existence of some relationships and distort the magnitude of others. This observation, however, is subject to several qualifications. Whether or not a functional health status measure will be appropriate for a particular study will depend upon the specific measure considered and the population being studied. For example, it has been shown that for a noninstitutionalized elderly population, while the Rosow-Breslau IADL scale is an excellent predictor of health services use and costs, the ADL scale is not, because of the relative insensitivity of the ADL scale in this population (over 90% of

respondents could perform all listed ADL activities).[31] On the other hand, for an institutionalized population, ADL has been shown to be highly predictive of resource use.[32] The particular aspect of health being investigated in a study also will affect the relative appropriateness of a functional measure or a perceived measure. As often observed,[23,33] health is a multi-faceted phenomenon, and the choice of one measure of physical health over another will be influenced by the characteristics of health which are being considered. Thus, if it is desired to focus specifically on episodic health, to incorporate the influence of emotional outlook and attitudes, or to adjust for respondents' perceptions of their health relative to others, a perceived measure may be preferred to a functional health status measure.

As we observed when describing the data for this study, the fact that utilization and expenditure estimates were based on six rather than twelve months of claims data for the period between the two health status surveys might be considered to limit the generalizability of our conclusions. However, we do not believe that this data limitation should affect our interpretation of the findings. Because of the way in which use and expenditure categories were constructed, subjects who experienced high volumes of utilization in the second six-month period (not included in our data), if incorrectly classified, would be placed into the lower use or expenditure categories. Had they been classified correctly, these individuals would probably have been in the moderate or high use categories. The net effect of this data limitation, therefore, was probably to yield slightly *lower* $R_i$'s for the low use/expenditure categories. Had twelve months of claims data been included rather than six, therefore, our observations and conclusions would likely have been strengthened rather than weakened.

## REFERENCES

1. German P: Measuring functional disability in the older population. *American Journal of Public Health.* 71:1197–1199, 1981.
2. Kane RA, Kane RL: *Assessing the elderly.* Lexington, MA: Lexington Books, 1981.
3. Ware JE, Brook RH, Davies AR, Lohr, KW: Choosing measures of health status for individuals in general populations. *American Journal of Public Health.* 71:620–625, 1981.
4. Stewart A, Ware JE, Brook RH: *Construction and Scoring of Aggregate Functional Status Indexes: Volume I.* Rand/R–2551–HHS. Santa Monica, CA: The Rand Corporation, 1981.
5. National Center for Health Statistics, Bonham, GS: Procedures and questionnaires of the National Medical Care Utilization and Expenditure Survey, *National Medical Care Utilization and Expenditure Survey.* Series A, Methodological Report No. 1 DHHS Pub No. 83–20001. Public Health Service, Washington: U.S. Government Printing Office, 1983.
6. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW: Studies of illness in the aged. *Journal of the American Medical Association.* 185:914–919, 1963.

7.  Rosow I, Breslau N: A Guttman health scale for the aged. *Journal of Gerontol.* 21:556–559, 1966.
8.  Weinberger M, Darnell JC, Tierney WM, et al.: Self-rated health as a predictor of hospital admission and nursing home placement in elderly public housing tenants *American Journal of Public Health* 76:457–459, 1986.
9.  Roos NP, Shapiro E: The Manitoba Longitudinal Study on Aging: Preliminary findings on health care utilization by the elderly. *Medical Care.* 19:644–657, 1981.
10. Eve SB: Older adult's use of health maintenance organizations: *Research on Aging.* 4:179–203, 1982.
11. Arling G: Interaction effects in a multivariate model of physician visits by older people. *Medical Care.* 23:361–371, 1985.
12. Soldo BJ, Manton KG: Health status and service needs of the oldest old: current patterns and future trends. *Milbank Memorial Fund Quarterly/Health and Society.* 63:287–319, 1985.
13. Branch LG, Katz S, Kniepmann K, Papsidero JA: A prospective study of functional status among community elders. *American Journal of Public Health.* 74:266–268, 1984.
14. Evashwick C, Rowe G, Diehr P, Branch L: Factors explaining the use of health services by the elderly. *Health Services Research.* 19:357–382, 1984.
15. Branch L, Jette A, Evashwick C, Polansky M, Rowe G, Dieler P: Toward understanding elder's health service utilization. *Journal of Community Health* 7:8–92, 1981.
16. Whitmore RW, Paul JE, Gibbs DA, Beebe JC: *Use of health status indicators in the calculation of the AAPCC.* Report prepared for the Health Care Financing Administration under Contract No. 500–81–0047, Research Triangle Institute, 1985.
17. Thomas JW, Lichtenstein RL: Including health status in Medicare's AAPCC capitation formula. *Medical Care.* 24:259–275, 1986.
18. Mossey JM, Shapiro E: Self-rated health: A predictor of mortality among the elderly. *American Journal of Public Health.* 72:800–808, 1982.
19. Katz S, Branch LG, Branson MH, Papsidero JA, Beck JC, Greer DS: Active life expectancy. *The New England Journal of Medicine.* 309:1218–1224, 1983.
20. Brook RH, Ware JE, Rogers WH et al.: Does free care improve adult's health? *New England Journal of Medicine.* 309:1426–1434, 1983.
21. Maddox GL, Douglass EB: Self-assessment of health: A longitudinal study of elderly subjects. *Journal of Health and Social Behavior.* 14:87–93, 1973.
22. Jette A, Branch LG: The Framingham disability study II: physical disability among the aging. *American Journal of Public Health* 71:1211–1216, 1981.
23. Liang J: The structure of self-reported physical health among aged adults. *Journal of Gerontology.* 41:248–260, 1986.
24. Thomas JW, Lichtenstein R, Wyszewianski L, Berki SE: *Final Report: A Health Status Measure for Adjusting the HMO Capitation Rates of Medicare Beneficiaries.* Prepared for the Health Care Financing Administration under Grant Number 18–P–98179/5–01. Ann Arbor, MI: Department of Medical Care Organization, School of Public Health, The University of Michigan, 1985.
25. Landis JR, Koch GG: The measurement of observer agreement for categorical data. *Biometrics.* 33:159–174, 1977.
26. Whitehurst GJ: Interrater agreement for journal manuscript review, *American Psychologist.* 39:22–28, 1984.
27. Bartko JJ: On various intraclass correlation reliability coefficients. *Psychological Bulletin.* 83:762–765, 1976.
28. Personal Communication from Richard Landis, Department of Biostatistics, University of Michigan.
29. Wolinsky FD, Coe RM, Miller DK, Prendergast JM: Correlates of change in subjective well-being among the elderly. *Journal of Community Health.* 10:93–107, 1985.
30. Haug M, Belgrave LL, Gratton B: Mental health and the elderly: Factors in stability and change over time. *Journal of Health and Social Behavior.* 25:100–115, 1984.
31. Thomas JW, Lichtenstein RL: Adjusting Medicare's health maintenance organization capitation formula for functional health status. *Health Care Financing Review.* 7:85–95, 1986.
32. Fries BE, Cooney L: Resource utilization groups: a patient classification system for long-term care. *Medical Care.* 23:110–122, 1985.
33. Donabedian A: *Aspects of Medical Care Administration.* Cambridge, Mass: Harvard University Press. 1973, pp. 138.